|  |
| --- |
| **머신러닝 및 딥러닝 앙상블 모델을 활용한 피싱사이트 탐지 연구** |

|  |
| --- |
| **이름1, ㅇㅇㅇ1, ㅇㅇㅇ1, ㅇㅇㅇ1**  **1**  **이메일---**  **영어 제목**  **영문 이름1, ooo, ooo, ooo** |

**Abstract**

본 논문은 웹사이트 내 피싱 사이트 탐지를 위한 머신러닝 및 딥러닝 앙상블 모델의 적용을 다룬다. 피싱 공격 수법의 다양화와 정교화로 인해 개인과 기업의 피해가 급증하고 있으며, 이에 대응하기 위한 정확하고 효율적인 탐지 기술의 필요성이 강조되고 있다. 본 연구에서는 피싱 사이트의 URL 및 콘텐츠 특징을 기반으로 머신러닝 모델 LightGBM(Light Gradient Boosting Machine)과 딥러닝 모델 CNN (Convolutional Neural Network)을 결합한 앙상블 기법을 활용하여 예측 성능을 분석한다. 이를 바탕으로, 악성 사이트 탐지 결과를 시각화하고, 웹 기반 시스템을 구축하여 사용자가 피싱 사이트를 보다 쉽게 식별할 수 있도록 지원한다.

I. 서론

피싱(Phishing)은 사용자의 민감한 정보를 도용하기 위한 사이버 공격으로, 이메일, 문자 메시지, 가짜 웹사이트 등을 통해 이루어진다. 피싱 사이트는 정상 사이트와 유사한 외관으로 사용자가 구별하기 어려우며, 그 결과 개인 기업에 심각한 경제적 피해를 초래하고 있다. 최근 피싱 공격이 정교해지면서 탐지 기술을 회피하기 위한 다양한 방법이 등장하고 있다. 특히, 블랙리스트 기반 탐지 방식은 한계가 있으며, 이에 따라 더 효과적인 탐지 기술의 필요성이 대두되고 있다.

본 연구에서는 피싱 사이트의 URL 특징, 보안 인증서, 콘텐츠 차이를 활용한 머신러닝 및 딥러닝 앙상블 모델을 제안한다. 이 모델은 피싱 사이트의 탐지의 정확도를 높이며, 보다 정교한 피싱 공격에 대응하는 새로운 탐지 방안을 제시하고자 한다.

II. 본론

2.1 데이터셋 구성

본 연구에서 사용된 데이터셋은 정상 URL과 피싱 URL을 포함하여 구성되었다. 학습용 데이터셋은 총 289,994개의 URL로 이루어져 있으며, 이 중 정상 URL 150,000개, 피싱 URL은 139,994개로 구성되어 있다. 검증 데이터셋은 정상 URL 50,000개, 피싱 URL 20,000개로 총 70,000개의 샘플을 포함한다. 이 데이터셋은 학습과 검증 과정에서 균형 있는 성능 평가를 목표로 하여 설계되었다.

2.2 피싱 사이트 탐지 모델

피싱 사이트 탐지 성능을 극대화하기 위해 머신러닝과 딥러닝 모델을 각각 적용하였다. 먼저, LightGBM과 CNN 모델을 독립적으로 학습시켜 피싱 사이트 분류 모델을 구축하였으며, 이후 두 모델의 예측 결과를 앙상블 기법으로 결합하여 최종적인 탐지 성능을 향상시켰다.

2.2.1 Light Gradient Boosting Machine(LightGBM)

LightGBM(Light Gradient Boosting Machine)은 속도와 정확도에서 뛰어난 성능을 발휘하는 머신러닝 알고리즘으로, 특히 대용량 데이터와 복잡한 구조의 데이터 분석에서 강점을 지닌다. 본 연구에서는 URL의 구조적 특징, 콘텐츠 특징 및 외부 쿼리 데이터를 바탕으로 피싱 사이트를 효과적으로 분류하기 위해 LightGBM을 활용하였다.

모델의 하이퍼파라미터 최적화는 Grid Search 기법을 통해 수행되었으며, 최적화된 하이퍼파라미터는 표 1과 같다. 최종 학습된 모델은 훈련과 검증 데이터셋에서 모두에서 우수한 성능을 보였으며, 모델 성능에 대한 평가는 표 2와 같다. 또한 검증 데이터셋을 기반으로 혼동 행렬(confusion matrix)을 사용하여 모델의 분류 성능을 시각적으로 평가한 결과, 피싱사이트(1)와 정상사이트(0)를 높은 정확도로 분류한 것을 확인할 수 있다.

추가적으로, LightGBM을 통해 도출된 특성 중요도(feature importance) 분석 결과, 총 198,323개의 특성 중 상위 20개의 중요한 특성은 그림 2와 같다. 가장 중요한 특성으로는 리다이렉션 여부, 특정 국가 도메인(kr, or, go) 및 하이퍼링크와 관련된 요소가 포함되었으며, 이는 리다이렉션과 특정 도메인의 사용이 피싱사이트 탐지에 있어 중요한 요소로 작용함을 보여준다.

|  |  |
| --- | --- |
| ***Parameter*** | ***Value*** |
| *learning\_rate* | 0.1 |
| *max\_depth* | -1 |
| *n\_estimators* | 100 |
| *num\_leaves* | 31 |

표 1. LightGBM 모델의 최적 하이퍼파라미터 설정

|  |  |  |
| --- | --- | --- |
|  | ***Test set*** | ***Valid set*** |
| *Accuracy* | 1.0 | 1.0 |
| *Precision* | 1.0 | 1.0 |
| *Recall* | 1.0 | 1.0 |
| *F1 Score* | 1.0 | 1.0 |

표 2. LightGBM 모델의 성능 지표

텍스트, 스크린샷, 도표, 폰트이(가) 표시된 사진

자동 생성된 설명

그림 1. LightGBM 모델의 피싱 사이트 탐지 결과에 대한 혼동 행렬

텍스트, 스크린샷, 도표, 그래프이(가) 표시된 사진

자동 생성된 설명

그림 2. LightGBM 모델을 통한 상위 20개 특성 중요도

2.2.2 Convolutional Neural Network(CNN)

Convolutional Neural Network(CNN)는 이미지 처리뿐만 아니라 텍스트 기반 데이터에서도 효과적으로 작동하는 딥러닝 모델이다. 본 연구에서는 URL 데이터를 CNN에 학습시켜 피싱 사이트를 탐지하였다. URL 데이터를 구성 요소(scheme, netloc, path)로 분리하여 토큰화한 후, 시퀀스 형태로 변환하고 등장 빈도가 낮은 단어를 제거하는 등의 전처리 과정을 거쳤다. 최종적으로 시퀀스 길이를 560으로 패딩 처리하였다.

CNN 모델은 첫 번째로 입력 시퀀스를 임베딩하는 Embedding 층으로 시작하여, 32개의 필터와 5의 커널 크기를 가진 Conv1D 층을 통해 1차원 컨볼루션을 수행하였다. 이후 GlobalMaxPooling1D 층을 통해 차원을 축소한 뒤, Dense 층에서 이진 분류를 수행하였으며, 활성화 함수로는 sigmoid를 사용하였다.

모델 학습은 binary\_crossentropy 손실 함수와 adam 옵티마이저를 사용하여 진행되었으며, 조기 종료와 모델 체크포인트 기법을 적용하여 최적 성능을 기록한 모델을 저장하였다. 학습은 총 10회의 에포크(epoch) 동안 진행되었으며, 검증 데이터셋을 통해 성능이 평가되었다. 모델 성능은 표 3에 제시되어 있으며, 각 층에서는 Dropout을 사용하여 과적합을 방지하고 모델의 일반화 성능을 향상시켰다.

|  |  |  |
| --- | --- | --- |
|  | ***Test set*** | ***Valid set*** |
| *Accuracy* | 0.98 | 0.93 |
| *Precision* | 0.99 | 0.98 |
| *Recall* | 0.97 | 0.79 |
| *F1 Score* | 0.98 | 0.87 |

표 3. CNN모델의 성능 지표

2.2.3 앙상블 기법

앙상블 모델은 두 모델의 예측 결과를 평균내어 최종 예측값을 산출하는 방식으로 이루어졌다. LightGBM과 CNN 모델의 예측 결과를 결합한 앙상블 예측값은 0.5를 기준으로 이진 클래스로 변환되었으며, 이를 통해 최종적인 피싱 사이트 탐지 결과를 도출하였다. 검증 데이터셋에 대한 앙상블 모델의 정확도는 0.99로 향상된 정확도를 기록하였다. 이를 통해 앙상블 기법이 피싱 사이트 탐지 성능을 향상시키는 데 효과적임을 확인하였다.

Ⅲ. 결론 및 향후 연구 방향

본 연구에서는 LightGBM 모델과 CNN 모델을 결합한 앙상블 기법을 적용하여 피싱 사이트 탐지 성능을 향상시켰다. LightGBM 모델은 URL의 구조적 특징을 바탕으로, CNN 모델은 URL 데이터를 바탕으로 각각 독립적으로 학습되었으며, 두 모델의 예측 결과를 결합함으로써 더욱 높은 정확도의 탐지 모델을 구축하였다.

LightGBM과 CNN 모델의 결합을 통해 피싱 사이트 탐지에서 다양한 데이터 특징을 효과적으로 분석할 수 있었으며, 이러한 접근 방식은 피싱 사이트의 복잡한 패턴을 탐지하는 데 유리한 것으로 확인되었다. URL의 세부 구성 요소와 콘텐츠 특징를 결합한 점에서 본 연구는 피싱 사이트 탐지 발전에 기여할 수 있을 것으로 기대된다.

향후 연구에서는 다양한 피싱 공격 수법에 대응하기 위해 보다 다양한 데이터셋을 활영하여 모델의 성능을 검증할 필요가 있다. 또한, 실시간 피싱 탐지를 가능하기 하기 위한 경량화된 모델 개발과 자동화된 특징 추출 기법을 연구함으로써, 실제 환경에서 피싱 사이트 탐지 시스템의 적용 가능성을 높이고, 웹 보안 강화틀에 실질적인 기여를 할 수 있을 것으로 기대된다.

참고문헌

[1]

[2]

[3]